



09-15-00

A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventorship..... Rui
Applicant..... Microsoft Corporation
Attorney's Docket No. .... MS1-610US
Title: Improved Image Retrieval Based on Relevance Feedback



TRANSMITTAL LETTER AND CERTIFICATE OF MAILING

To: Commissioner of Patents and Trademarks,
Washington, D.C. 20231

From: Allan T. Sponseller (Tel. 509-324-9256; Fax 509-323-8979)
Lee & Hayes, PLLC
421 W. Riverside Avenue, Suite 500
Spokane, WA 99201

The following enumerated items accompany this transmittal letter and are being submitted for the
matter identified in the above caption.

- 1. Specification--title page, plus 42 pages, including 50 claims and Abstract
2. Transmittal letter including Certificate of Express Mailing
3. 5 Sheets Formal Drawings (Figs. 1-5)
4. Return Post Card

Large Entity Status [x] Small Entity Status [ ]

Date: Sept. 13, 2000

By: [Signature]
Allan T. Sponseller
Reg. No. 38,318

CERTIFICATE OF MAILING

I hereby certify that the items listed above as enclosed are being deposited with the U.S. Postal
Service as either first class mail, or Express Mail if the blank for Express Mail No. is completed below, in
an envelope addressed to The Commissioner of Patents and Trademarks, Washington, D.C. 20231, on the
below-indicated date. Any Express Mail No. has also been marked on the listed items.

Express Mail No. (if applicable) EL624352666

Date: 9-13-00

By: [Signature]
Lori A. Vierra

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

**Improved Image Retrieval Based On Relevance  
Feedback**

Inventor(s):

Yong Rui

ATTORNEY'S DOCKET NO. MS1-610US

1 **RELATED APPLICATIONS**

2 This application claims the benefit of U.S. Provisional Application No.  
3 60/153,730, filed September 13, 1999, entitled "MPEG-7 Enhanced Multimedia  
4 Access" to Yong Rui, Jonathan Grudin, Anoop Gupta, and Liwei He, which is  
5 hereby incorporated by reference.  
6

7 **TECHNICAL FIELD**

8 This invention relates to image storage and retrieval, and more particularly  
9 to retrieving images based on relevance feedback.  
10

11 **BACKGROUND OF THE INVENTION**

12 Computer technology has advanced greatly in recent years, allowing the  
13 uses for computers to similarly grow. One such use is the storage of images.  
14 Databases of images that are accessible to computers are constantly expanding and  
15 cover a wide range of areas, including stock images that are made commercially  
16 available, images of art collections (e.g., by museums), etc. However, as the  
17 number of such images being stored has increased, so too has the difficulty in  
18 managing the retrieval of such images. Often times it is difficult for a user to  
19 search databases of such images to identify selected ones of the thousands of  
20 images that are available.

21 One difficulty in searching image databases is the manner in which images  
22 are stored versus the manner in which people think about and view images. It is  
23 possible to extract various low-level features regarding images, such as the color  
24 of particular portions of an image and shapes identified within an image, and make  
25 those features available to an image search engine. However, people don't tend to

1 think of images using such low-level features. For example, a user that desires to  
2 retrieve images of brown dogs would typically not be willing and/or able to input  
3 search parameters identifying the necessary color codes and particular areas  
4 including those color codes, plus whatever low-level shape features are necessary  
5 to describe the shape of a dog in order to retrieve those images. Thus, there is  
6 currently a significant gap between the capabilities provided by image search  
7 engines and the usability desired by people using such engines.

8 One solution is to provide a text-based description of images. In  
9 accordance with this solution, images are individually and manually categorized  
10 by people, and various descriptive words for each image are added to a database.  
11 For example, a picture of a brown dog licking a small boy's face may include key  
12 words such as dog, brown, child, laugh, humor, etc. There are, however, problems  
13 with this solution. One such problem is that it requires manual categorization – an  
14 individual(s) must take the time to look at a picture, decide which key words to  
15 include for the picture, and record those key words. Another problem is that such  
16 a process is subjective. People tend to view images in different ways, viewing  
17 shapes, colors, and other features differently. With such a manual process, the key  
18 words will be skewed towards the way the individual cataloging the images views  
19 the images, and thus different from the way many other people will view the  
20 images.

21 The invention described below addresses these disadvantages, providing for  
22 improved image retrieval based on relevance feedback.

## 24 **SUMMARY OF THE INVENTION**

25 Improved image retrieval based on relevance feedback is described herein.

1 According to one aspect, a hierarchical (per-feature) approach is used in  
2 comparing images. Multiple query vectors are generated for an initial image by  
3 extracting multiple low-level features from the initial image. When determining  
4 how closely a particular image in an image collection matches that initial image, a  
5 distance is calculated between the query vectors and corresponding low-level  
6 feature vectors extracted from the particular image. Once these individual  
7 distances are calculated, they are combined to generate an overall distance that  
8 represents how closely the two images match.

9 According to another aspect, when a set of potentially relevant images are  
10 presented to a user, the user is given the opportunity to provide feedback regarding  
11 the relevancy of the individual images in the set. This relevancy feedback is then  
12 used to generate a new set of potentially relevant images for presentation to the  
13 user. The relevancy feedback is used to influence the generation of the query  
14 vector, influence the weights assigned to individual distances between query  
15 vectors and feature vectors when generating an overall distance, and to influence  
16 the determination of the distances between the query vectors and the feature  
17 vectors.

18 According to another aspect, the calculation of a distance between a query  
19 vector and a feature vector involves the use of a matrix to weight the individual  
20 vector elements. The type of matrix used varies dynamically based on the number  
21 of images for which feedback has been received from the user and the number of  
22 feature elements in the feature vector. If the number of images for which feedback  
23 has been received is less than the number of feature elements, then a diagonal  
24 matrix is used (which assigns weights to the individual vector elements in the  
25 distance calculation). However, if the number of images for which feedback has

1 been received equals or exceeds the number of feature elements, then a full matrix  
2 is used (which transforms the low-level features of the query vector and the  
3 feature vector to a higher level feature space, as well as assigns weights to the  
4 individual transformed elements in the distance calculation).

## 6 **BRIEF DESCRIPTION OF THE DRAWINGS**

7 The present invention is illustrated by way of example and not limitation in  
8 the figures of the accompanying drawings. The same numbers are used  
9 throughout the figures to reference like components and/or features.

10 Fig. 1 is a block diagram illustrating an exemplary network environment  
11 such as may be used in accordance with certain embodiments of the invention.

12 Fig. 2 illustrates an example of a suitable operating environment in which  
13 the invention may be implemented.

14 Fig. 3 is a block diagram illustrating an exemplary image retrieval  
15 architecture in accordance with certain embodiments of the invention.

16 Fig. 4 is a flowchart illustrating an exemplary process, from the perspective  
17 of a client, for using relevance feedback to retrieve images.

18 Fig. 5 is a flowchart illustrating an exemplary process, from the perspective  
19 of an image server, for using relevance feedback to retrieve images.

## 21 **DETAILED DESCRIPTION**

22 Fig. 1 is a block diagram illustrating an exemplary network environment  
23 such as may be used in accordance with certain embodiments of the invention. In  
24 the network environment 100 of Fig. 1, an image server 102 is coupled to one or  
25 more image collections 104. Each image collection stores one or more images of

1 a wide variety of types. In one implementation, the images are still images,  
2 although it is to be appreciated that other types of images can also be used with the  
3 invention. For example, each frame of moving video can be treated as a single  
4 still image. Image collections 104 may be coupled directly to image server 102,  
5 incorporated into image server 102, or alternatively indirectly coupled to image  
6 server 102 such as via a network 106.

7 Also coupled to image server 102 is one or more client devices 108. Client  
8 devices 108 may be coupled to image server 102 directly or alternatively indirectly  
9 (such as via network 106). Image server 102 acts as an interface between clients  
10 108 and image collections 104. Image server 102 allows clients 108 to retrieve  
11 images from image collections 104 and render those images. Users of clients 108  
12 can then input relevance feedback, which is returned to image server 102 and used  
13 to refine the image retrieval process, as discussed in more detail below.

14 Network 106 represents any of a wide variety of wired and/or wireless  
15 networks, including public and/or private networks (such as the Internet, local area  
16 networks (LANs), wide area networks (WANs), etc.). A client 108, image server  
17 102, or image collection 104 can be coupled to network 106 in any of a wide  
18 variety of conventional manners, such as wired or wireless modems, direct  
19 network connections, etc.

20 Communication among devices coupled to network 106 can be  
21 accomplished using one or more protocols. In one implementation, network 106  
22 includes the Internet. Information is communicated among devices coupled to the  
23 Internet using, for example, the well-known Hypertext Transfer Protocol (HTTP),  
24 although other protocols (either public and/or proprietary) could alternatively be  
25 used.

Fig. 2 illustrates an example of a suitable operating environment in which the invention may be implemented. The illustrated operating environment is only one example of a suitable operating environment and is not intended to suggest any limitation as to the scope of use or functionality of the invention. Other well known computing systems, environments, and/or configurations that may be suitable for use with the invention include, but are not limited to, personal computers, server computers, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, programmable consumer electronics (e.g., digital video recorders), gaming consoles, cellular telephones, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

Fig. 2 shows a general example of a computer 142 that can be used in accordance with the invention. Computer 142 is shown as an example of a computer that can perform the functions of client 108 or server 102 of Fig. 1. Computer 142 includes one or more processors or processing units 144, a system memory 146, and a bus 148 that couples various system components including the system memory 146 to processors 144.

The bus 148 represents one or more of any of several types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. The system memory 146 includes read only memory (ROM) 150 and random access memory (RAM) 152. A basic input/output system (BIOS) 154, containing the basic routines that help to transfer information between elements within computer 142, such as during start-up, is stored in ROM 150. Computer 142 further includes a hard disk drive 156 for reading from and writing



to a hard disk, not shown, connected to bus 148 via a hard disk drive interface 157 (e.g., a SCSI, ATA, or other type of interface); a magnetic disk drive 158 for reading from and writing to a removable magnetic disk 160, connected to bus 148 via a magnetic disk drive interface 161; and an optical disk drive 162 for reading from and/or writing to a removable optical disk 164 such as a CD ROM, DVD, or other optical media, connected to bus 148 via an optical drive interface 165. The drives and their associated computer-readable media provide nonvolatile storage of computer readable instructions, data structures, program modules and other data for computer 142. Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 160 and a removable optical disk 164, it will be appreciated by those skilled in the art that other types of computer readable media which can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, random access memories (RAMs), read only memories (ROM), and the like, may also be used in the exemplary operating environment.

A number of program modules may be stored on the hard disk, magnetic disk 160, optical disk 164, ROM 150, or RAM 152, including an operating system 170, one or more application programs 172, other program modules 174, and program data 176. A user may enter commands and information into computer 142 through input devices such as keyboard 178 and pointing device 180. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner, or the like. These and other input devices are connected to the processing unit 144 through an interface 168 that is coupled to the system bus (e.g., a serial port interface, a parallel port interface, a universal serial bus (USB) interface, etc.). A monitor 184 or other type of display device is also connected to

1 the system bus 148 via an interface, such as a video adapter 186. In addition to the  
2 monitor, personal computers typically include other peripheral output devices (not  
3 shown) such as speakers and printers.

4 Computer 142 operates in a networked environment using logical  
5 connections to one or more remote computers, such as a remote computer 188.  
6 The remote computer 188 may be another personal computer, a server, a router, a  
7 network PC, a peer device or other common network node, and typically includes  
8 many or all of the elements described above relative to computer 142, although  
9 only a memory storage device 190 has been illustrated in Fig. 2. The logical  
10 connections depicted in Fig. 2 include a local area network (LAN) 192 and a wide  
11 area network (WAN) 194. Such networking environments are commonplace in  
12 offices, enterprise-wide computer networks, intranets, and the Internet. In certain  
13 embodiments of the invention, computer 142 executes an Internet Web browser  
14 program (which may optionally be integrated into the operating system 170) such  
15 as the "Internet Explorer" Web browser manufactured and distributed by  
16 Microsoft Corporation of Redmond, Washington.

17 When used in a LAN networking environment, computer 142 is connected  
18 to the local network 192 through a network interface or adapter 196. When used  
19 in a WAN networking environment, computer 142 typically includes a modem 198  
20 or other means for establishing communications over the wide area network 194,  
21 such as the Internet. The modem 198, which may be internal or external, is  
22 connected to the system bus 148 via a serial port interface 168. In a networked  
23 environment, program modules depicted relative to the personal computer 142, or  
24 portions thereof, may be stored in the remote memory storage device. It will be  
25

1 appreciated that the network connections shown are exemplary and other means of  
2 establishing a communications link between the computers may be used.

3 Computer 142 also includes a broadcast tuner 200. Broadcast tuner 200  
4 receives broadcast signals either directly (e.g., analog or digital cable  
5 transmissions fed directly into tuner 200) or via a reception device (e.g., via  
6 antenna 110 or satellite dish 114 of Fig. 1).

7 Computer 142 typically includes at least some form of computer readable  
8 media. Computer readable media can be any available media that can be accessed  
9 by computer 142. By way of example, and not limitation, computer readable  
10 media may comprise computer storage media and communication media.  
11 Computer storage media includes volatile and nonvolatile, removable and non-  
12 removable media implemented in any method or technology for storage of  
13 information such as computer readable instructions, data structures, program  
14 modules or other data. Computer storage media includes, but is not limited to,  
15 RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM,  
16 digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic  
17 tape, magnetic disk storage or other magnetic storage devices, or any other media  
18 which can be used to store the desired information and which can be accessed by  
19 computer 142. Communication media typically embodies computer readable  
20 instructions, data structures, program modules or other data in a modulated data  
21 signal such as a carrier wave or other transport mechanism and includes any  
22 information delivery media. The term "modulated data signal" means a signal that  
23 has one or more of its characteristics set or changed in such a manner as to encode  
24 information in the signal. By way of example, and not limitation, communication  
25 media includes wired media such as wired network or direct-wired connection,

1 and wireless media such as acoustic, RF, infrared and other wireless media.  
2 Combinations of any of the above should also be included within the scope of  
3 computer readable media.

4 The invention has been described in part in the general context of  
5 computer-executable instructions, such as program modules, executed by one or  
6 more computers or other devices. Generally, program modules include routines,  
7 programs, objects, components, data structures, etc. that perform particular tasks  
8 or implement particular abstract data types. Typically the functionality of the  
9 program modules may be combined or distributed as desired in various  
10 embodiments.

11 For purposes of illustration, programs and other executable program  
12 components such as the operating system are illustrated herein as discrete blocks,  
13 although it is recognized that such programs and components reside at various  
14 times in different storage components of the computer, and are executed by the  
15 data processor(s) of the computer.

16 Alternatively, the invention may be implemented in hardware or a  
17 combination of hardware, software, and/or firmware. For example, one or more  
18 application specific integrated circuits (ASICs) could be designed or programmed  
19 to carry out the invention.

20 Fig. 3 is a block diagram illustrating an exemplary image retrieval  
21 architecture in accordance with certain embodiments of the invention. The image  
22 retrieval architecture 220 illustrated in Fig. 3 is implemented, for example, in an  
23 image server 102 of Fig. 1. Architecture 220 includes a query vector generator  
24 222, a comparator 224, multiple images 226 and corresponding low-level image  
25 features 228, and an image retriever 230.

Multiple low-level features are extracted for each image 226. These features are described as being extracted prior to the image retrieval process discussed herein, although the features could alternatively be extracted during the image retrieval process. Each feature is a vector (referred to as a feature vector) that includes multiple feature elements. The number of feature elements in a feature vector can vary on a per-feature basis.

Low-level image features 228 can include any of a wide variety of conventional features, such as: color moment features, color histogram features, wavelet texture features, Fourier descriptor features, water-fill edge features, etc. In one implementation, low-level features 228 include three features: a color moments feature, a wavelet based texture feature, and a water-fill edge feature. The color moments feature is a 6-element vector obtained by extracting the mean and standard deviation from three color channels in the HSV (hue, saturation, value) color space. The wavelet based texture feature is a 10-element vector obtained by a wavelet filter bank decomposing the image into 10 de-correlated sub-bands, with each sub-band capturing the characteristics of a certain scale and orientation of the original image. The standard deviation of the wavelet coefficients for each sub-band is extracted, and these standard deviations used as the elements of the feature vector. The water-fill edge feature is an 18-element vector that is obtained by extracting 18 different elements from the edge maps: the maximum filling time and associated fork count, the maximum fork count and associated filing time, the filling time histogram for each of seven bins (ranges of values), and the fork count histogram for each of seven bins. For additional information regarding the water-fill edge feature can be found in Xiang Sean Zhou, Yong Rui, and Thomas S. Huang, "Water-Filling: A Novel Way for Image

1 Structural Feature Extraction", Proc. of IEEE International Conference on Image  
2 Processing, Kobe, Japan, October 1999, which is hereby incorporated by  
3 reference.

4 Low-level image features 228 can be stored and made accessible in any of a  
5 wide variety of formats. In one implementation, the low-level features 228 are  
6 generated and stored in accordance with the MPEG-7 (Moving Pictures Expert  
7 Group) format. The MPEG-7 format standardizes a set of Descriptors (Ds) that  
8 can be used to describe various types of multimedia content, as well as a set of  
9 Description Schemes (DSs) to specify the structure of the Ds and their  
10 relationship. In MPEG-7, the individual features 228 are each described as one or  
11 more Descriptors, and the combination of features is described as a Description  
12 Scheme.

13 During the image retrieval process, search criteria in the form of an initial  
14 image selection 232 is input to query vector generator 222. The initial image  
15 selection 232 can be in any of a wide variety of forms. For example, the initial  
16 image may be an image chosen from images 226 in accordance with some other  
17 retrieval process (e.g., based on a descriptive keyword search), the image may be  
18 an image that belongs to the user and is not included in images 226, etc. The  
19 initial selection 232 may or may not include low-level features for the image. If  
20 low-level features that will be used by comparator 224 are not included, then those  
21 low-level features are generated by query vector generator 222 based on initial  
22 selection 232 in a conventional manner. Note that these may be the same features  
23 as low-level image features 228, or alternatively a subset of the features 228.  
24 However, if the low-level features are already included, then query vector  
25 generator 222 need not generate them. Regardless of whether generator 222

1 generates the low-level features for initial image selection 232, these low-level  
2 features are output by query vector generator 222 as query vectors 234.

3       Comparator 224 performs an image comparison based on the low-level  
4 image features 228 and the query vectors 234. This comparison includes possibly  
5 mapping both the low-level image features 228 and the query vectors 234 to a  
6 higher level feature space and determining how closely the transformed (mapped)  
7 features and query vectors match. An identification 236 of a set of potentially  
8 relevant images is then output by comparator 224 to image retriever 230. The  
9 potentially relevant images are those images that comparator 224 determines have  
10 low-level image features 228 most closely matching the query vectors. Retriever  
11 230 obtains the identified images from images 226 and returns those images to the  
12 requestor (e.g., a client 108 of Fig. 1) as potentially relevant images 238.

13       A user is then able to provide relevance feedback 240 to query vector  
14 generator 222. In one implementation, each of the potentially relevant images 238  
15 is displayed to the user at a client device along with a corresponding graphical  
16 "degree of relevance" slider. The user is able to slide the slider along a slide bar  
17 ranging from, for example, "Not Relevant" to "Highly Relevant". Each location  
18 along the slide bar that the slider can be positioned at by the user has a  
19 corresponding value that is returned to the generator 222 and comparator 224 and  
20 incorporated into their processes as discussed in more detail below. In one  
21 implementation, if the user provides no feedback, then a default relevancy  
22 feedback is assigned to the image (e.g., equivalent to "no opinion"). Alternatively,  
23 other user interface mechanisms may be used to receive user feedback, such as  
24 radio buttons corresponding to multiple different relevancy feedbacks (e.g., Highly  
25

1 Relevant, Relevant, No Opinion, Irrelevant, and Highly Irrelevant), verbal  
2 feedback (e.g., via speech recognition), etc.

3 The relevance feedback is used by query vector generator 222 to generate a  
4 new query vector and comparator 224 to identify a new set of potentially relevant  
5 images. The user relevance feedback 240 can be numeric values that are directly  
6 used by generator 222 and comparator 224, such as: an integer or real value from  
7 zero to ten; an integer or real value from negative five to positive five; values  
8 corresponding to highly relevant, somewhat relevant, no opinion, somewhat  
9 irrelevant, and highly irrelevant of 7, 3, 0, -3, and -7, respectively. Alternatively,  
10 the user relevance feedback 240 can be an indication in some other format (e.g.,  
11 the text or encoding of "Highly Relevant") and converted to a useable numeric  
12 value by generator 222, comparator 224, and/or another component (not  
13 illustrated).

14 The second set of potentially relevant images displayed to the user is  
15 determined by comparator 224 incorporating the relevance feedback 240 received  
16 from the user into the comparison process. This process can be repeated any  
17 number of times, with the feedback provided each time being used to further refine  
18 the image retrieval process.

19 Note that the components illustrated in architecture 220 may be distributed  
20 across multiple devices. For example, low-level features 228 may be stored  
21 locally at image server 102 of Fig. 1 (e.g., on a local hard drive) while images 226  
22 may be stored at one or more remote locations (e.g., accessed via network 106).

23 The image retrieval process discussed herein refers to several different  
24 types of matrixes, including diagonal matrixes, full matrixes, and the identity  
25 matrix. A diagonal matrix refers to a matrix that can have any value along the



diagonal, where the diagonal of a matrix  $B$  are the elements of the matrix at positions  $B_{jj}$ , and values not along the diagonal are zero. The identity matrix is a special case of the diagonal matrix where the elements of the matrix along the diagonal all have the value of one and all other elements in the matrix have a value of zero. A full matrix is a matrix in which any element can have any value. These different types of matrixes are well-known to those skilled in the art, and thus will not be discussed further except as they pertain to the present invention.

The specific manner in which query vectors are generated, comparisons are made, and relevance feedback is incorporated into both of these processes will now be described. It is to be appreciated that these specific manners described are only examples of the processes and that various modifications can be made to the these descriptions.

Each single image of the images 226 has multiple ( $I$ ) corresponding low-level features in the features 228. As used herein,  $\vec{x}_{mi}$  refers to the  $i^{th}$  feature vector of the  $m^{th}$  image, so:

$$\vec{x}_{mi} = [x_{mi1}, \dots, x_{mik}, \dots, x_{miK_i}]$$

where  $K_i$  is the length of the feature vector  $\vec{x}_{mi}$ .

A query vector is generated as necessary for each of the low-level feature spaces. The query vector is initially generated by extracting the low-level feature elements in each of the feature spaces from the initial selection 232. The query vector can be subsequently modified by the relevance feedback 240, as discussed in more detail below. The query vector in a feature space  $i$  is:

$$\vec{q}_i = [q_{i1}, \dots, q_{ik}, \dots, q_{iK_i}]$$

To compare the query vector ( $\vec{q}_i$ ) and a corresponding feature vector of an image  $m$  ( $\vec{x}_{mi}$ ), the distance between the two vectors is determined. A wide variety of different distance metrics can be used, and in one implementation the generalized Euclidean distance is used. The generalized Euclidean distance between the two vectors, referred to as  $g_{mi}$ , is calculated as follows:

$$g_{mi} = (\vec{q}_i - \vec{x}_{mi})^T W_i (\vec{q}_i - \vec{x}_{mi})$$

where  $W_i$  is a matrix that both optionally transforms the low-level feature space into a higher level feature space and then assigns weights to each feature element in the higher level feature space. When sufficient data is available to perform the transformation, the low-level feature space is transformed into a higher level feature space that better models user desired high-level concepts.

The matrix  $W_i$  can be decomposed as follows:

$$W_i = P_i^T \Lambda_i P_i$$

where  $P_i$  is an orthonormal matrix consisting of the eigen vectors of  $W_i$ , and  $\Lambda_i$  is a diagonal matrix whose diagonal elements are the eigen values of  $W_i$ . Thus, the calculation to determine the distance  $g_{mi}$  can be rewritten as:

$$g_{mi} = (P_i(\vec{q}_i - \vec{x}_{mi}))^T \Lambda_i (P_i(\vec{q}_i - \vec{x}_{mi}))$$

where the low-level feature space is transformed into the higher level feature space by the mapping matrix  $P_i$  and then weights are assigned to the feature elements of the new feature space by the weighting matrix  $\Lambda_i$ .

However, in some situations there may be insufficient data to reliably perform the transformation into the higher level feature space. In such situations, the matrix  $W_i$  is simply the weighting matrix  $\Lambda_i$ , so  $g_{mi}$  can be rewritten as:

$$g_{mi} = (\vec{q}_i - \vec{x}_{mi})^T \Lambda_i (\vec{q}_i - \vec{x}_{mi}).$$

Typically, each of multiple ( $I$ ) low-level feature vectors of images in the database is compared to a corresponding query vector and the individual distances between these vectors determined. Once all of the  $I$  low-level feature vectors have been compared to the corresponding query vectors and distances determined, these distances are combined to generate an overall distance  $d_m$ , which is defined as follows:

$$d_m = U(g_{mi})$$

where  $U()$  is a function that combines the individual distances  $g_{mi}$  to form the overall distance  $d_m$ . Thus, a hierarchical approach is taken to determining how closely two images match: first individual distances between the feature vectors and the query vectors are determined, and then these individual distances are combined.

The function  $U()$  can be any of a variety of different combinatorial functions. In one implementation, the function  $U()$  is a weighted summation of the individual distances, resulting in:

$$d_m = \sum_{i=1}^I u_i [(\vec{q}_i - \vec{x}_{mi})^T W_i (\vec{q}_i - \vec{x}_{mi})]$$

The feature vectors of the individual images ( $\vec{x}_{mi}$ ) are known (they are features 228). The additional values needed to solve for the overall distance  $d_m$  are: the

weights ( $u_i$ ) of each individual feature distance, the query vector ( $\vec{q}_i$ ) for each feature, and the transformation matrix ( $W_i$ ) for each feature. For the first comparison (before any relevance feedback 240 is received), each query vector ( $\vec{q}_i$ ) is simply the corresponding extracted feature elements of the initial selection 232, the weights ( $u_i$ ) of each individual distance are the same (e.g., a value of  $1/I$ , where  $I$  is the number of features used), and each transformation matrix ( $W_i$ ) is the identity matrix. The determination of these individual values based on relevance feedback is discussed in more detail below.

Alternatively, the generalized Euclidean distance could also be used to compute  $d_m$ , as follows:

$$d_m = \vec{g}_{mi}^T U \vec{g}_{mi}$$

where  $U$  is an ( $I \times I$ ) full matrix.

The overall distance  $d_m$  is thus calculated for each image 226. Alternatively, the overall distance  $d_m$  may be calculated for only a subset of images 226. Which subset of images 226 to use can be identified in any of a variety of manners, such as using well-known multi-dimensional indexing techniques (e.g., R-tree or R\*-tree).

A number of images 226 having the smallest distance  $d_m$  are then selected as potentially relevant images to be presented to a user. The number of images 226 can vary, and in one implementation is determined empirically based on both the size of display devices typically being used to view the images and the size of the images themselves. In one implementation, twenty images are returned as potentially relevant.

User relevance feedback 240 identifies degrees of relevance for one or more of the potentially relevant images 238 (that is, a value indicating how relevant each of one or more of the images 238 is). A user may indicate that only selected ones of the images 238 are relevant, and user relevance feedback 240 identify degrees of relevance for only those selected images. Alternatively, user relevance feedback 240 may identify degrees of relevance for all images 238, such as by assigning a default value to those images for which the user did not assign a relevancy. These default values (and corresponding image features) can then be ignored by query vector generator 222 and comparator 224 (e.g., dropped from relevance feedback 240), or alternatively treated as user input feedback and used by vector generator 222 and comparator 224 when generating new values.

Once relevance feedback 240 is received, query vector generator 222 generates new query vectors 234. The new query vectors are referred to as  $\vec{q}_i^*$ , and are defined as follows:

$$\vec{q}_i^* = \frac{\vec{\pi}^T X_i}{\sum_{n=1}^N \pi_n}$$

where  $N$  represents the number of potentially relevant images for which the user input relevance feedback (e.g., non-default relevance values were returned), which can be less than the number of potentially relevant images that were displayed to the user ( $N$  may also be referred to as the number of training samples);  $\pi_n$  represents the degree of relevance of image  $n$  as indicated by the relevance feedback from the user (that is, a degree of relevance value associated with the relevance indicated by the user),  $\vec{\pi}^T$  represents a  $(1 \times N)$  vector of the individual  $\pi_n$  values, and  $X_i$  represents a training sample matrix for feature  $I$  that is

1 obtained by stacking the  $N$  training vectors ( $\vec{x}_{ni}$ ) into a matrix, and resulting in an  
2 ( $N \times K_i$ ) matrix.

3 Alternatively,  $N$  (both here and elsewhere in this discussion) may  
4 represent the number of potentially relevant images for which relevance feedback  
5 was received regardless of the source (e.g., including both user-input feedback and  
6 default relevance values).

7 The process of presenting potentially relevant images to a user and  
8 receiving relevance feedback for at least portions of that set of potentially relevant  
9 images can be repeated multiple times. The results of each set of feedback can be  
10 saved and used for determining subsequent query vectors (as well as the weights  
11 ( $u_i$ ) of each individual distance and each transformation matrix ( $W_i$ )) in the  
12 process, or alternatively only a certain number of preceding sets of feedback may  
13 be used. For example, if three sets of twenty images each are presented to a user  
14 and relevance feedback returned for each image of the three sets, then to generate  
15 the fourth set the feedback from all sixty images may be used. Alternatively, only  
16 the feedback from the most recent set of twenty images may be used (or the two  
17 most recent sets, etc.).

18 Comparator 224 also receives relevance feedback 240 and uses relevance  
19 feedback 240 to generate a new value for  $W_i$ , which is referred to as  $W_i^*$ . The  
20 value of  $W_i^*$  is either a full matrix or a diagonal matrix. When the number of  
21 potentially relevant images for which the user input relevance feedback ( $N$ ) is less  
22 than the length of the feature vector ( $K_i$ ), the value of  $W_i^*$  as a full matrix cannot  
23 be calculated (and is difficult to reliably estimate, if possible at all). Thus, in  
24 situations where  $N < K_i$ ,  $W_i^*$  is a diagonal matrix; otherwise  $W_i^*$  is a full matrix.

To generate the full matrix,  $W_i^*$  is calculated as follows:

$$W_i^* = (\det(C_i))^{\frac{1}{K_i}} C_i^{-1}$$

where  $\det(C_i)$  is the matrix determinant of  $C_i$ , and  $C_i$  is the  $(K_i \times K_i)$  weighted covariance matrix of  $X_i$ . In other words,

$$C_{i_{rs}} = \frac{\sum_{n=1}^N \pi_n (x_{nir} - q_{ir})(x_{nis} - q_{is})}{\sum_{n=1}^N \pi_n}$$

where  $r$  is the row index of the matrix  $C_i$  and ranges from 1 to  $K_i$ ,  $s$  is the column index of the matrix  $C_i$  and ranges from 1 to  $K_i$ ,  $N$  represents the number of potentially relevant images for which the user input relevance feedback,  $\pi_n$  represents the degree of relevance of image  $n$ ,  $x_{nir}$  refers to the  $r^{\text{th}}$  element of the feature vector for feature  $i$  of image  $n$ ,  $q_{ir}$  refers to the  $r^{\text{th}}$  element of the query vector for feature  $i$ ,  $x_{nis}$  refers to the  $s^{\text{th}}$  element of the feature vector for feature  $i$  of the  $n^{\text{th}}$  image, and  $q_{is}$  refers to the  $s^{\text{th}}$  element of the query vector for feature  $i$ .

To generate the diagonal matrix, each diagonal element of the matrix is calculated as follows:

$$w_{i_{kk}} = \frac{1}{\sigma_{ik}}$$

where  $w_{i_{kk}}$  is the  $kk^{\text{th}}$  element of matrix  $W_i$  and  $\sigma_{ik}$  is the standard deviation of the sequence of  $x_{ik}$ 's, and where each  $x_{ik}$  is the  $k^{\text{th}}$  element of feature  $i$ .

It should be noted that the determination of whether  $W_i$  is to be a full matrix or a diagonal matrix is done on a per-image basis as well as a per-feature

1 basis for each image. Thus, depending on the length of each feature vector,  $W_i$   
2 may be different types of matrixes for different features.

3 It should also be noted that in situations where  $W_i$  is a diagonal matrix, the  
4 distance ( $g_{mi}$ ) between a query vector ( $\vec{q}_i$ ) and a feature vector ( $\vec{x}_{mi}$ ) is based on  
5 weighting the feature elements but not transforming the feature elements to a  
6 higher level feature space. This is because there is an insufficient number of  
7 training samples to reliably perform the transformation. However, in situations  
8 where  $W_i$  is a full matrix, the distance ( $g_{mi}$ ) between a query vector ( $\vec{q}_i$ ) and a  
9 feature vector ( $\vec{x}_{mi}$ ) is based on both transforming the low-level features to a  
10 higher level feature space and weighting the transformed feature elements.

11 Once relevance feedback 240 is received, comparator 224 also generates a  
12 new value for  $u_i$ , which is referred to as  $u_i^*$ , and is calculated as follows:

$$13 \quad u_i^* = \sum_{j=1}^I \sqrt{\frac{f_j}{f_i}}$$

14  
15 where

$$16 \quad f_i = \sum_{n=1}^N \pi_n g_{ni}$$

17  
18 where  $N$  represents the number of potentially relevant images for which the user  
19 input relevance feedback,  $\pi_n$  represents the degree of relevance of image  $n$ , and  
20  $g_{ni}$  ( $g_{mi}$  as discussed above) represents the distance between the previous query  
21 vector ( $\vec{q}_i$ ) and the feature vector ( $\vec{x}_{ni}$ ).  
22

23 Fig. 4 is a flowchart illustrating an exemplary process, from the perspective  
24 of a client, for using relevance feedback to retrieve images. The process of Fig. 4  
25



1 is carried out by a client 108 of Fig. 1, and can be implemented in software. Fig. 4  
2 is discussed with reference to components in Figs. 1 and 3.

3 First, initial search criteria (e.g., an image) is entered by the user (act 260).  
4 The initial search criteria is used by image server 102 to identify potentially  
5 relevant images 238 which are received (from server 102) and rendered at client  
6 108 (act 262) as the initial search results. The client then receives an indication  
7 from the user as to whether the search results are satisfactory. This indication can  
8 be direct (e.g., selection of an on-screen button indicating that the results are  
9 satisfactory or to stop the retrieval process) or indirect (e.g., input of relevance  
10 feedback indicating that one or more of the images is not relevant). If the search  
11 results are satisfactory, then the process ends (act 266).

12 However, if the search results are not satisfactory, then the relevance of the  
13 search results is identified (act 268). The relevance of one or more images in the  
14 search results is identified by user feedback (e.g., user selection of one of multiple  
15 options indicating how relevant the image is). A new search request that includes  
16 the relevance feedback regarding the search results is then submitted to server 102  
17 (act 270). In response to the search request, the server 102 generates new search  
18 results (based in part on the relevance feedback), which are received by client 108  
19 (act 272). The process then returns to act 264, allowing for additional user  
20 relevance feedback as needed.

21 Fig. 5 is a flowchart illustrating an exemplary process, from the perspective  
22 of an image server, for using relevance feedback to retrieve images. The process  
23 of Fig. 5 is carried out by an image server 102 of Fig. 1, and can be implemented  
24 in software. Fig. 5 is discussed with reference to components in Figs. 1 and 3.

1 To begin the image retrieval process, search criteria are received by image  
2 server 102 (act 282) as initial selection 232, in response to which generator 222  
3 generates multiple query vectors (act 284). Comparator 224 then maps the low-  
4 level feature vectors of images in image collection 104 to a higher level feature  
5 vector for each image and compares the higher level feature vectors to the query  
6 vector (act 286). The images that most closely match the query vectors (based on  
7 the comparison in act 286) are then identified (act 288), and forwarded to the  
8 requesting client 108 (act 290). Alternatively, in some situations the mapping to  
9 the higher level feature space may not occur, and the comparison and  
10 identification may be performed based on the low-level feature space.

11 Server 102 then receives user feedback from the requesting client 108  
12 regarding the relevance of one or more of the identified images (act 292). Upon  
13 receipt of this relevance feedback, generator 222 generates a new query vector  
14 based in part on the relevance feedback and comparator 224 uses the relevance  
15 feedback to generate a new transformation matrix and new feature distance  
16 weights (act 294). The process then returns to act 286, where the new mapping  
17 parameters and new query vector are used to identify new images for forwarding  
18 to the client.

## 19 20 **Conclusion**

21 Although the description above uses language that is specific to structural  
22 features and/or methodological acts, it is to be understood that the invention  
23 defined in the appended claims is not limited to the specific features or acts  
24 described. Rather, the specific features and acts are disclosed as exemplary forms  
25 of implementing the invention.

## CLAIMS

1. One or more computer readable media having stored thereon a plurality of instructions that, when executed by one or more processors, causes the one or more processors to perform acts including:

receiving an initial image selection;

generating a plurality of query vectors by extracting, for each query vector, one of a plurality of low-level features from the initial image selection;

selecting a set of potentially relevant images based at least in part on distances between the plurality of query vectors and a plurality of feature vectors corresponding to low-level features of a plurality of images;

receiving feedback regarding the relevance of one or more images of the set of potentially relevant images;

generating a new plurality of query vectors based at least in part on the feedback;

generating a weighting of feature elements based at least in part on the feedback; and

selecting a new set of potentially relevant images based at least in part on both the weighting of feature elements and distances between the new plurality of query vectors and the plurality of feature vectors.

1           2. One or more computer readable media as recited in claim 1, wherein  
2 the selecting a new set of potentially relevant images comprises using a matrix in  
3 determining the distance between one of the new plurality of query vectors and  
4 one of the plurality of feature vectors, and further comprising dynamically  
5 selecting the matrix based on both a number of images in the set of potentially  
6 relevant images for which relevance feedback was input and a number of feature  
7 elements in the one feature vector.

8  
9           3. One or more computer readable media as recited in claim 2, wherein  
10 the dynamically selecting comprises using a diagonal matrix if the number of  
11 images in the set of potentially relevant images for which relevance feedback was  
12 input is less than the number of feature elements in the one feature vector, and  
13 otherwise using a full matrix.

14  
15           4. One or more computer readable media as recited in claim 2, wherein  
16 the dynamically selecting comprises:

17           if the number of images in the set of potentially relevant images for which  
18 relevance feedback was input is not less than the number of feature elements in the  
19 one feature vector, then using one matrix that transforms the query vector and the  
20 one feature vector to a higher-level feature space and then using another matrix  
21 that assigns a weight to each element of the transformed query vector and the  
22 transformed feature vector; and

23           if the number of images in the set of potentially relevant images is less than  
24 the number of feature elements in the one feature vector, then using a matrix that  
25 assigns a weight to each element of the query vector and the one feature vector.

1  
2 5. One or more computer readable media as recited in claim 2, wherein  
3  $X$  represents an image matrix that is generated by stacking  $N$  feature vectors  
4 corresponding to the set of potentially relevant images for which relevance  
5 feedback was received and resulting in an  $(N \times K)$  matrix,  $C$  represents a  
6 weighted covariance matrix of  $X$ ,  $\det(C)$  represents the matrix determinant of  $C$ ,  
7 and the matrix comprises a full matrix ( $W^*$ ) that is generated as follows:

8 
$$W^* = (\det(C))^{\frac{1}{K}} C^{-1}.$$
  
9  
10

11 6. One or more computer readable media as recited in claim 2, wherein  
12  $w_{kk}$  represents the  $kk^{\text{th}}$  element of matrix  $W$ ,  $x_k$  represents the  $k^{\text{th}}$  feature element,  
13  $\sigma_k$  represents the standard deviation of the sequence of  $x_k$ 's, the matrix comprises  
14 a diagonal matrix with each diagonal element ( $w_{kk}$ ) being generated as follows:

15 
$$w_{kk} = \frac{1}{\sigma_k}$$
  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

7. One or more computer readable media as recited in claim 1, wherein  $N$  represents the number of images in the set of potentially relevant images for which relevance feedback has been received,  $\pi_n$  represents the relevance of image  $n$  in the set of images,  $\pi^{\rightarrow T}$  represents a transposition of a vector generated by concatenating the individual  $\pi_n$  values, and  $X$  represents an image matrix that is generated by stacking  $N$  training vectors corresponding to the set of potentially relevant images into a matrix, and wherein each new query vector ( $\vec{q}$ ) of the new plurality of query vectors is generated as follows:

$$\vec{q} = \frac{\pi^{\rightarrow T} X}{\sum_{n=1}^N \pi_n}.$$

8. One or more computer readable media as recited in claim 1, wherein  $f_i$  represents a summation, over the images in the set of potentially relevant images, of a product of a relevance of the image and a distance between the query vector and the feature vector, and wherein the selecting a new set of potentially relevant images comprises combining, for each image, a weighted distance between the plurality of query vectors and the plurality of feature vectors, and wherein the weight ( $u_i$ ) for each of a plurality ( $I$ ) of distances between a query vector and a corresponding feature vector is calculated as:

$$u_i = \sum_{j=1}^I \sqrt{\frac{f_j}{f_i}}.$$

9. One or more computer readable media as recited in claim 1, wherein the receiving feedback comprises receiving feedback from a user.

1           **10.**    One or more computer readable media as recited in claim 1, wherein  
2 the low-level features include: a color moments feature, a wavelet based texture  
3 feature, and a water-fill edge feature.  
4

5           **11.**    A method of selecting between two types of matrixes to be used to  
6 weight, based on relevance feedback, a plurality of feature elements for image  
7 retrieval, the method comprising:

8                selecting one of the two types of matrixes based on both a number of  
9 previously retrieved relevant images and a length of a feature vector including the  
10 plurality of feature elements.  
11

12           **12.**    A method as recited in claim 11, wherein the selecting comprises  
13 selecting one of the two types of matrixes based on both a number of previously  
14 retrieved potentially relevant images which were identified by a user as being  
15 relevant, and the length of the feature vector including the plurality of feature  
16 elements.  
17

18           **13.**    A method as recited in claim 11, wherein the plurality of feature  
19 elements are all elements of the same feature.  
20

21           **14.**    A method as recited in claim 11, wherein the selecting comprises  
22 using a first type of matrix if the number of retrieved relevant images is less than  
23 the length of the feature vector, and otherwise using a second type of matrix.  
24  
25

1           **15.**    A method as recited in claim 14, wherein the first type of matrix  
2 comprises a diagonal matrix and wherein the second type of matrix comprises a  
3 full matrix.

4  
5           **16.**    A method as recited in claim 11, wherein the selecting comprises  
6 using a first type of matrix if the length of the feature vector exceeds the number  
7 of retrieved relevant images by at least a threshold amount, and otherwise using a  
8 second type of matrix.

9  
10          **17.**    A method as recited in claim 16, wherein the first type of matrix  
11 comprises a full matrix and the second type of matrix comprises a diagonal matrix.

12  
13          **18.**    One or more computer readable media including a computer  
14 program that is executable by a processor to perform the method recited in claim  
15 11.

16  
17          **19.**    One or more computer readable media having stored thereon a  
18 plurality of instructions that, when executed by one or more processors, causes the  
19 one or more processors to perform acts including:

20                comparing, for each of a plurality of images, a plurality of feature elements  
21 from a query vector to a plurality of feature elements from a feature vector  
22 corresponding to the image;

23                identifying a number of potentially relevant images based on the  
24 comparing;



1 receiving user feedback regarding the relevancy of one or more of the  
2 potentially relevant images;

3 re-comparing, for each of the plurality of images, the plurality of feature  
4 elements from the query vector to the plurality of feature elements from the  
5 feature vector, including using a matrix to compare the feature elements and  
6 dynamically selecting a type of matrix to use based on both the user feedback and  
7 the number of the plurality of feature elements;

8 identifying a new set of potentially relevant images based on the re-  
9 comparing; and

10 presenting the new set of potentially relevant images to the user.  
11

12 **20.** One or more computer readable media as recited in claim 19,  
13 wherein the re-comparing comprises dynamically selecting the type of matrix to  
14 use based on both a number of the potentially relevant images for which user  
15 feedback has been received and the number of the plurality of feature elements.  
16

17 **21.** One or more computer readable media as recited in claim 19,  
18 wherein the dynamically weighting comprises using a first type of matrix if the  
19 number of retrieved relevant images is less than the length of the feature vector,  
20 and otherwise using a second type of matrix.  
21

22 **22.** One or more computer readable media as recited in 21, wherein the  
23 first type of matrix comprises a diagonal matrix and the second type of matrix  
24 comprises a full matrix.  
25

1       **23.**    A method comprising:  
2       generating a query vector corresponding to a feature of one image;  
3       identifying a feature vector corresponding to the feature of another image;  
4       identifying a number of training samples for which relevance feedback has  
5       been received;

6       if the number of training samples either equals or exceeds a threshold  
7       amount, then determining a distance between the query vector and the feature  
8       vector including transforming the query vector and the feature vector to a higher-  
9       level feature space and then assigning a weight to each element of the transformed  
10      query vector and the transformed feature vector; and

11      if the number of training samples does not exceed the threshold amount,  
12      then determining the distance between the query vector and the feature vector  
13      including assigning a weight to each element of the query vector and the feature  
14      vector.

15  
16      **24.**    A method as recited in claim 23, wherein the feature vector includes  
17      a plurality of feature elements and wherein the threshold amount comprises the  
18      number of feature elements in the feature vector.

19  
20      **25.**    A method as recited in claim 23, wherein if the number of training  
21      samples either equals or exceeds the threshold amount, then determining the  
22      distance ( $g$ ), where  $P$  is a mapping matrix,  $\vec{q}$  is the query vector,  $\vec{x}$  is the feature  
23      vector, and  $\Lambda$  is a weighting matrix, as:

24      
$$g = (P(\vec{q} - \vec{x}))^T \Lambda (P(\vec{q} - \vec{x})).$$
  
25

1       **26.**     A method as recited in claim 23, wherein if the number of training  
2 samples does not exceed the threshold amount, then determining the distance ( $g$ ),  
3 where  $\vec{q}$  is the query vector,  $\vec{x}$  is the feature vector, and  $\Lambda$  is a weighting matrix,  
4 as:

$$5 \quad g = (\vec{q} - \vec{x})^T \Lambda (\vec{q} - \vec{x}).$$

6  
7       **27.**     A method as recited in claim 23, further comprising:  
8       repeating the generating, identifying of the feature vector, identifying of the  
9 number of training samples, and the determining for each of a plurality of features;  
10 and  
11       identifying how closely the image and the other image match each other by  
12 combining the distances between the query vectors and the feature vectors for the  
13 plurality of features.

14  
15       **28.**     A method as recited in claim 27, wherein the identifying comprises  
16 calculating a weighted summation of each of the individual distances for each of  
17 the plurality of features.

18  
19       **29.**     One or more computer readable media including a computer  
20 program that is executable by a processor to perform the method recited in claim  
21 23.

22  
23       **30.**     A system comprising:  
24       a query vector generator to generate a query vector corresponding to a  
25 feature of one image;

1 a comparator, coupled to the query vector generator, to,

2 identify a feature vector corresponding to the feature of another  
3 image,

4 identify a number of training samples for which relevance feedback  
5 has been received,

6 if the number of training samples either equals or exceeds a  
7 threshold amount, then to determine a distance between the query vector  
8 and the feature vector including transforming the query vector and the  
9 feature vector to a higher-level feature space and then assigning a weight to  
10 each element of the transformed query vector and the transformed feature  
11 vector, and

12 if the number of training samples does not exceed the threshold  
13 amount, then to determine the distance between the query vector and the  
14 feature vector including assigning a weight to each element of the query  
15 vector and the feature vector.

16  
17 **31.** A method comprising:

18 for one of a plurality of images and each of a plurality of features,

19 generating, based on the set of search criteria, a query vector for the  
20 feature,

21 identifying a feature vector, corresponding to the image, for the  
22 feature, and

23 determining how closely the feature vector matches the query vector;  
24  
25

1 determining how closely the image matches the set of search criteria based  
2 on how closely, for the plurality of features, the feature vectors match the query  
3 vectors.

4  
5 **32.** A method as recited in claim 31, wherein generating the query  
6 vector comprises generating the query vector based at least in part on user  
7 relevance feedback regarding how relevant images previously displayed to a user  
8 were.

9  
10 **33.** A method as recited in claim 31, wherein identifying the feature  
11 vector comprises:

12 identifying a low-level feature vector corresponding to the feature; and

13 mapping the low-level feature vector to a higher level feature space.

14  
15 **34.** A method as recited in claim 33, wherein the identifying the feature  
16 vector further comprises incorporating, into the mapping, relevance feedback.

17  
18 **35.** A method as recited in claim 31, wherein the initial search criteria  
19 comprises an image.  
20  
21  
22  
23  
24  
25

1       **36.**    A method as recited in claim 31, wherein the determining how  
2 closely the feature vector matches the query vector comprises determining a  
3 distance between the feature vector and the query vector, and wherein the  
4 determining how closely the image matches the set of search criteria comprises  
5 calculating a weighted summation of each of the individual distances between the  
6 feature vectors and the query vectors.

7  
8       **37.**    A method as recited in claim 36, wherein the calculating a weighted  
9 summation comprises calculating the weighted summation based at least in part on  
10 user relevance feedback regarding how relevant images previously displayed to a  
11 user were.

12  
13       **38.**    One or more computer readable media including a computer  
14 program that is executable by a processor to perform the method recited in claim  
15 31.

16  
17       **39.**    One or more computer readable media having stored thereon a  
18 plurality of instructions that, when executed by one or more processors, causes the  
19 one or more processors to perform acts including:

20       identifying a plurality of query vectors for one image, each query vector  
21 corresponding to one of a plurality of features;

22       identifying a plurality of feature vectors for the other image, each feature  
23 vector corresponding to one of the plurality of features;

24       for each feature, determining a distance between the corresponding query  
25 vector and the corresponding feature vector; and

1 combining the distances to generate a value representing an overall distance  
2 between the two images.

3  
4 **40.** One or more computer readable media as recited in claim 39,  
5 wherein the identifying the plurality of query vectors comprises extracting the  
6 plurality of query vectors from the image.

7  
8 **41.** One or more computer readable media as recited in claim 39,  
9 wherein the identifying the plurality of query vectors comprises generating the  
10 plurality of query vectors based at least in part on user relevance feedback  
11 regarding how relevant images previously displayed to a user were.

12  
13 **42.** One or more computer readable media as recited in claim 39,  
14 wherein the determining the distance between the corresponding query vector and  
15 the corresponding feature vector includes incorporating, into the determining, user  
16 relevance feedback regarding how relevant images previously displayed to a user  
17 were.

18  
19 **43.** One or more computer readable media as recited in claim 39,  
20 wherein the combining the distances comprises calculating a weighted summation  
21 of each of the individual distances between the feature vectors and the query  
22 vectors.

1       **44.** One or more computer readable media as recited in claim 43,  
2 wherein the calculating a weighted summation comprises calculating the weighted  
3 summation based at least in part on user relevance feedback regarding how  
4 relevant images previously displayed to a user were.

5  
6       **45.** A system comprising:  
7 a query vector generator to identify a query vector, for each of a plurality of  
8 features, corresponding to one image;  
9 a comparator, coupled to the query vector generator, to,  
10 identify, for each of a plurality of features of another image, a  
11 feature vector for the feature,  
12 determine, for each of the plurality of features, how closely the  
13 feature vector matches the query vector, and  
14 determine how closely the image matches the other based on how  
15 closely, for the plurality of features, the feature vectors match the query  
16 vectors.

17  
18       **46.** A method of generating a query vector to compare to a feature  
19 vector of another image, the method comprising:  
20 receiving feedback regarding the relevance of each image of a set of  
21 images;  
22 wherein  $N$  represents the number of images in the set of images for which  
23 user relevance feedback has been received,  $\pi_n$  represents the relevance of image  $n$   
24 in the set of images,  $\pi^{\rightarrow T}$  represents a transposition of a vector generated by  
25 concatenating the individual  $\pi_n$  values, and  $X$  represents an image matrix that is



generated by stacking  $N$  training vectors corresponding to the set of images into a matrix; and

generating a query vector ( $\vec{q}$ ) corresponding to one of a plurality of features as follows:

$$\vec{q} = \frac{\vec{\pi}^T X}{\sum_{n=1}^N \pi_n}.$$

**47.** One or more computer readable media including a computer program that is executable by a processor to perform the method recited in claim 46.

**48.** A method of generating a weight to apply to distances between query vectors and feature vectors when combining the distances, the method comprising:

receiving feedback regarding the relevance of each image of a set of images;

wherein  $f_i$  represents a summation, over the images in the set of images, of a product of a relevance of the image and a distance between the query vector and the feature vector; and

generating a weight ( $u_i$ ) for each of a plurality ( $I$ ) of distances between a query vector corresponding to one of a plurality ( $I$ ) of features and a feature vector corresponding to the one of the plurality ( $I$ ) of features as:

$$u_i = \sum_{j=1}^I \sqrt{\frac{f_j}{f_i}}.$$

1  
2       **49.** One or more computer readable media including a computer  
3 program that is executable by a processor to perform the method recited in claim  
4 48.

5  
6       **50.** A system comprising:  
7 a client device;  
8 a collection of a plurality of images;  
9 an image server, coupled to the client device and the collection of a  
10 plurality of images, the image server to receive image retrieval requests from the  
11 client device and to,

12           receive an initial image selection from the client device,

13           generate a plurality of query vectors by extracting, for each query  
14 vector, one of a plurality of low-level features from the initial image  
15 selection,

16           select a set of potentially relevant images based at least in part on  
17 distances between the plurality of query vectors and a plurality of feature  
18 vectors corresponding to low-level features of a plurality of images,

19           receive feedback regarding the relevance of one or more images of  
20 the set of potentially relevant images,

21           generate a new plurality of query vectors based at least in part on the  
22 feedback,

23           generate a weighting of feature elements based at least in part on the  
24 feedback, and  
25

1 select a new set of potentially relevant images based at least in part  
2 on both the weighting of feature elements and distances between the new  
3 plurality of query vectors and the plurality of feature vectors.  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

## **ABSTRACT**

An improved image retrieval process based on relevance feedback uses a hierarchical (per-feature) approach in comparing images. Multiple query vectors are generated for an initial image by extracting multiple low-level features from the initial image. When determining how closely a particular image in an image collection matches the initial image, a distance is calculated between the query vectors and corresponding low-level feature vectors extracted from the particular image. Once these individual distances are calculated, they are combined to generate an overall distance that represents how closely the two images match. According to other aspects, relevancy feedback received regarding previously retrieved images is used during the query vector generation and the distance determination to influence which images are subsequently retrieved.

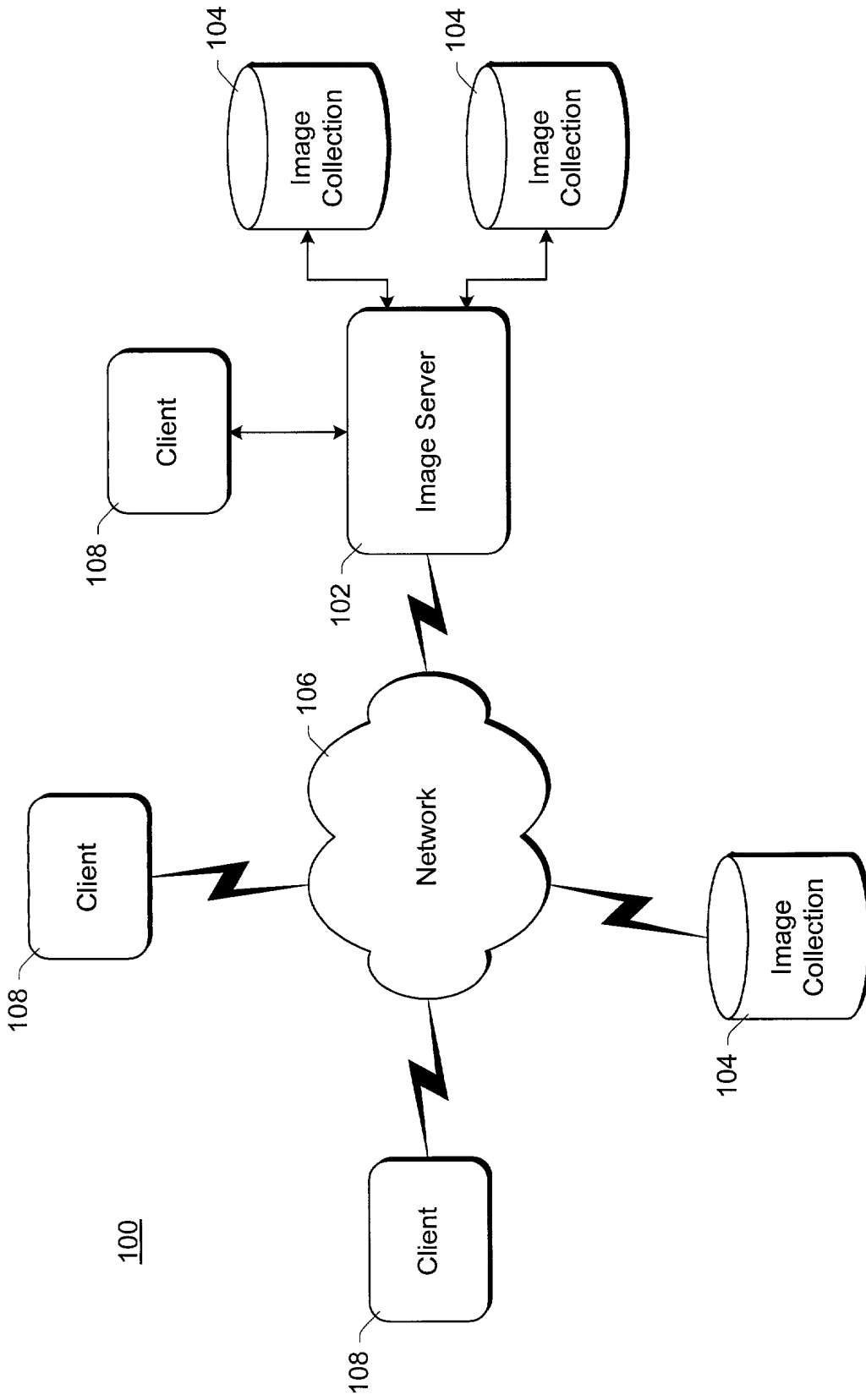
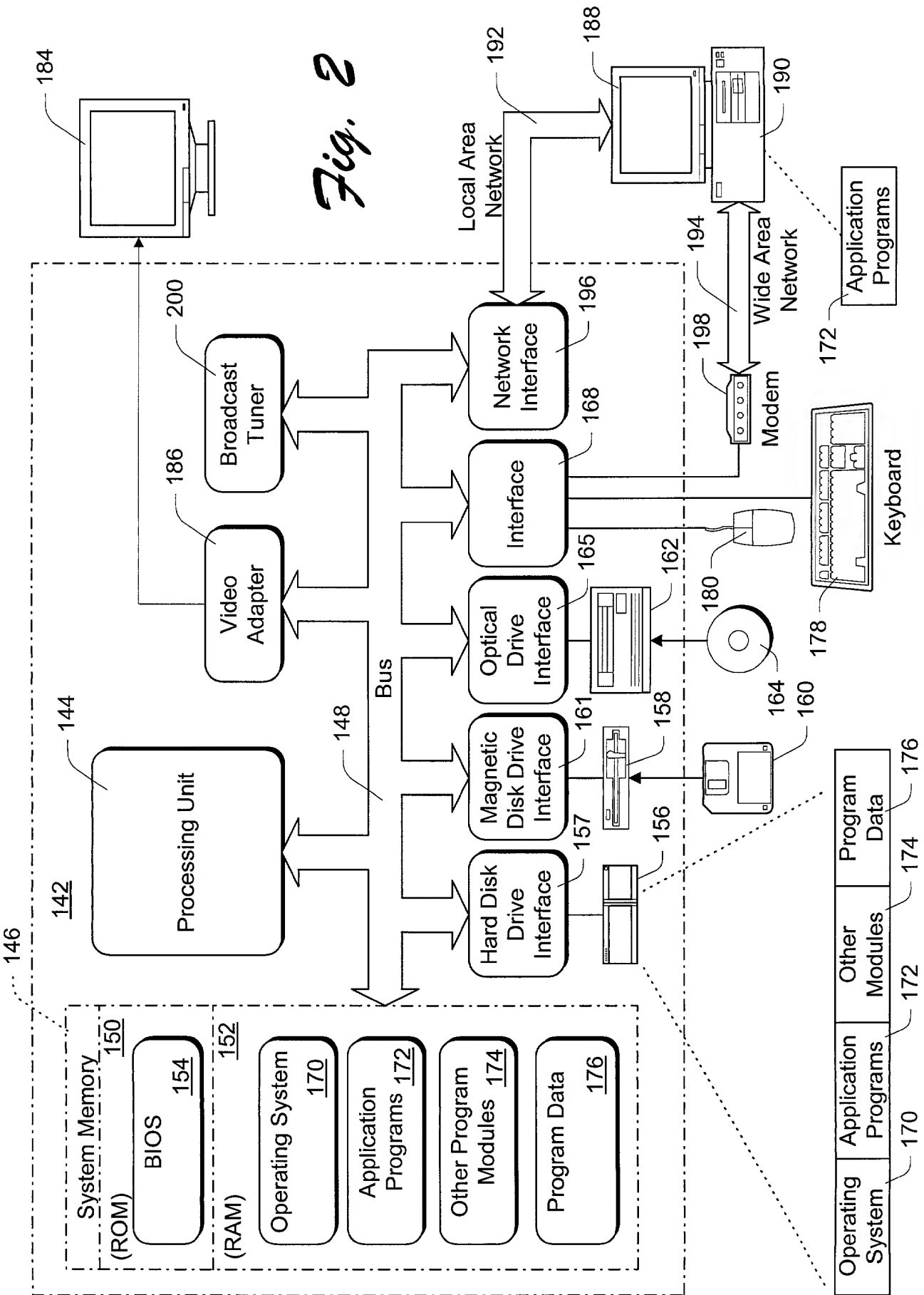


Fig. 1

Fig. 2



220

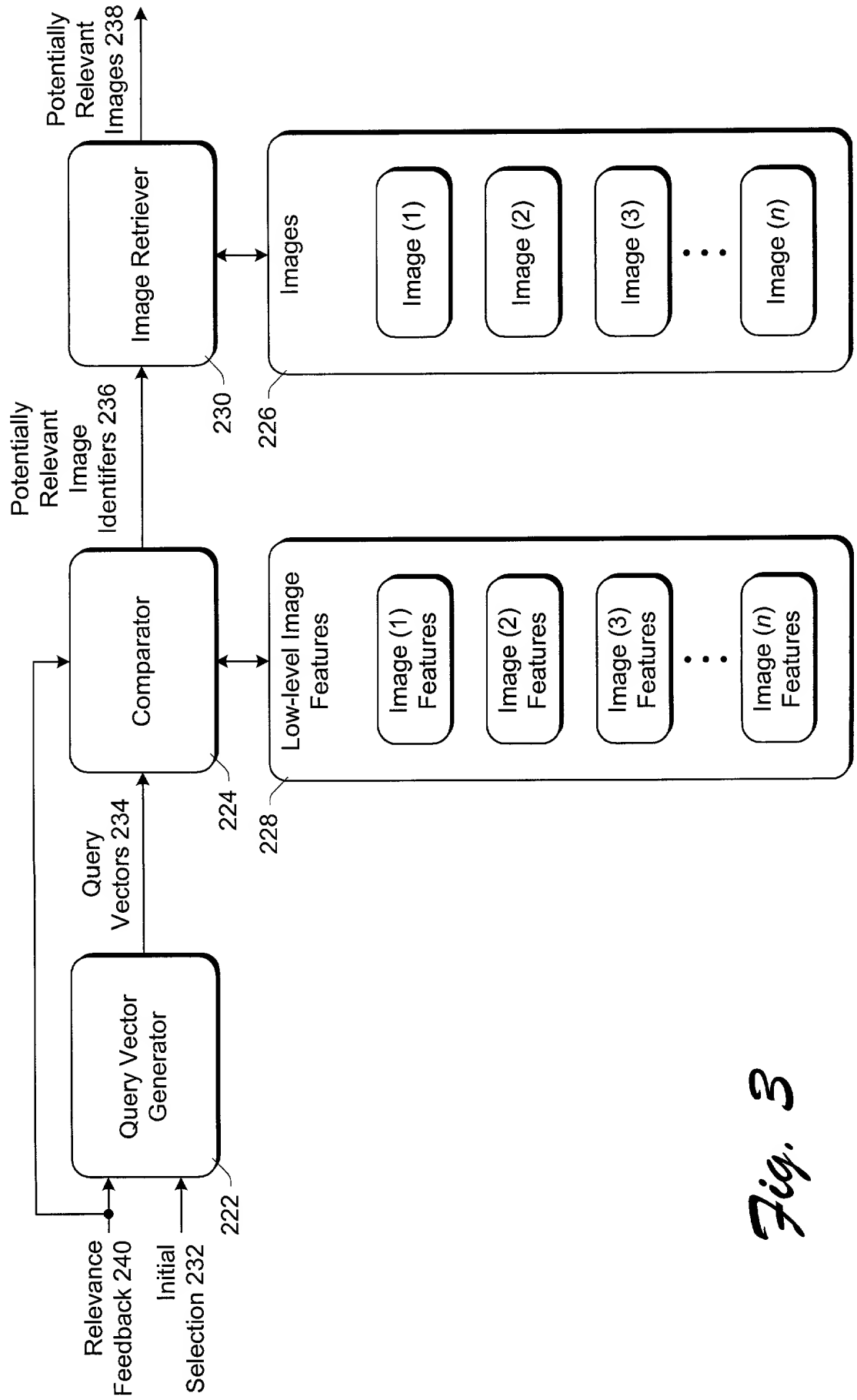
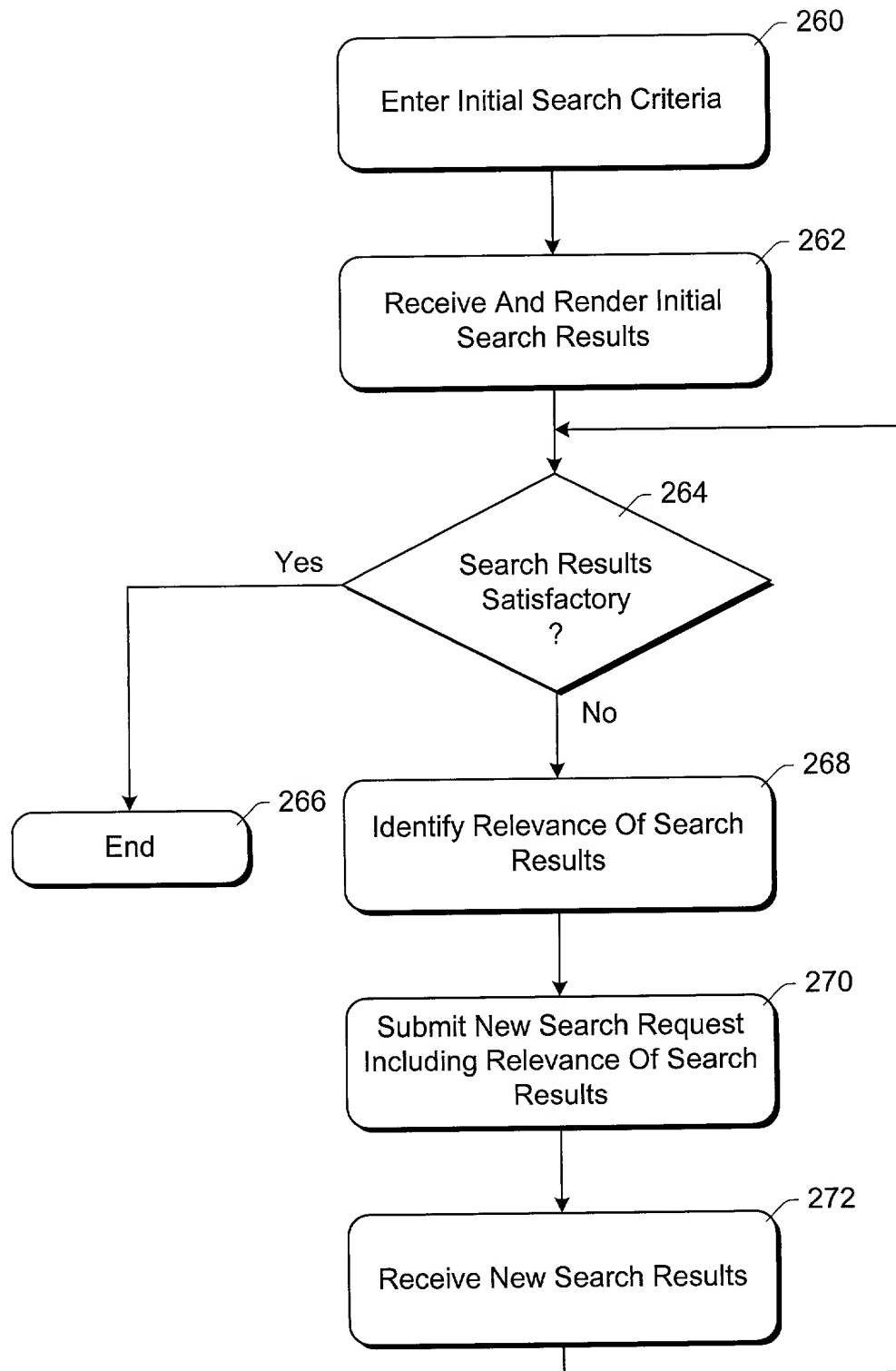
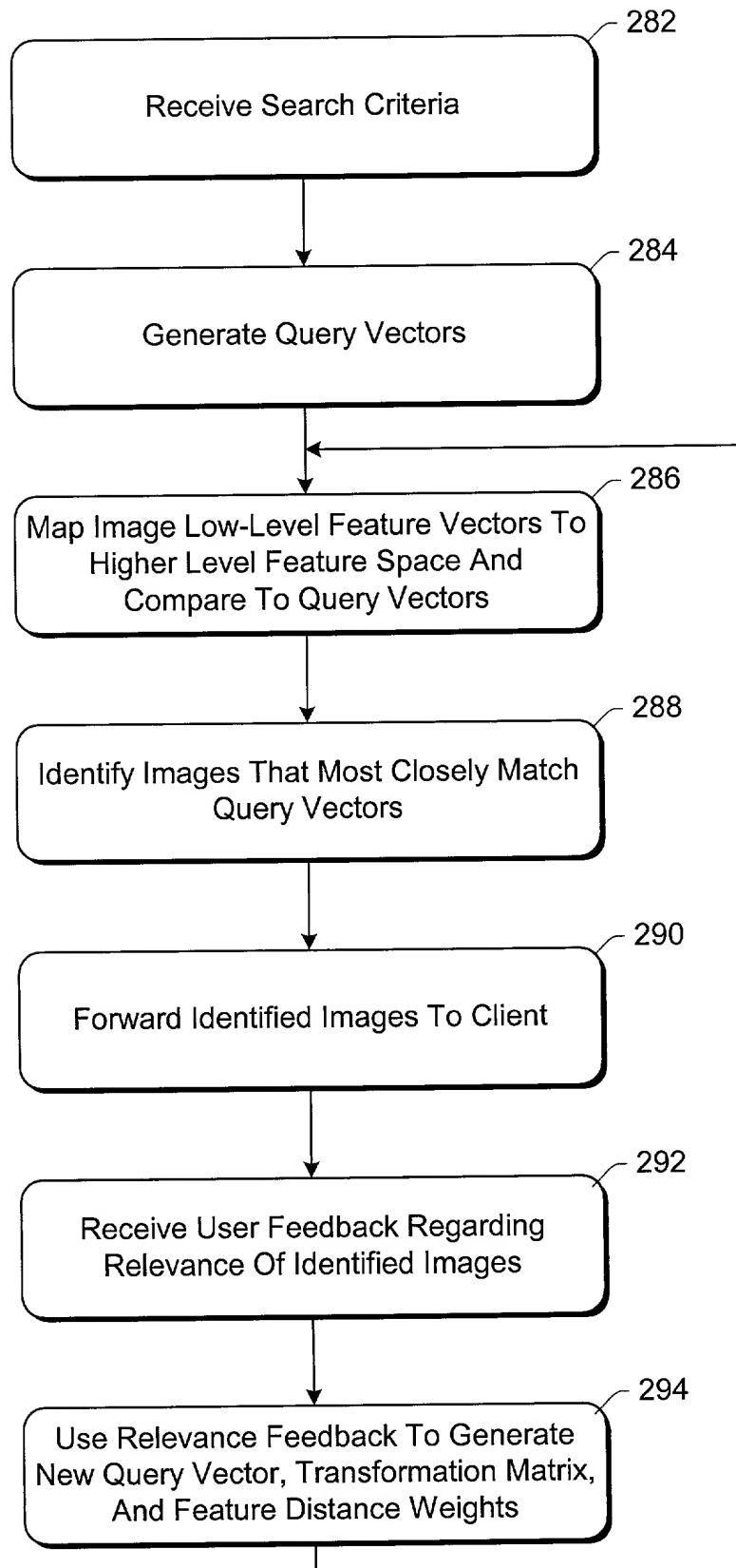


Fig. 3

*Fig. 4*



*Fig. 5*